

Scanned

Department Copy

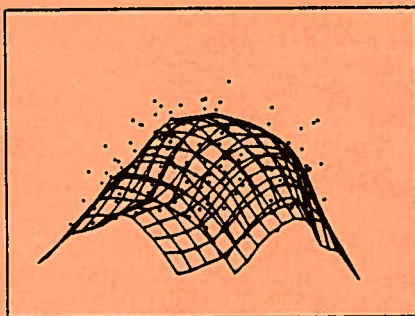
# **BOOTSTRAP CONFIDENCE INTERVALS**

*Robert Tibshirani*

**Technical Report No. 3**

**October 1984**

**Laboratory for  
Computational  
Statistics**



**Department of Statistics  
Stanford University**

This document and the material and data contained therein, was developed under sponsorship of the United States Government. Neither the United States nor the Department of Energy, nor the Office of Naval Research, nor the U.S. Army Research Office, nor the Leland Stanford Junior University, nor their employees, nor their respective contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any liability or responsibility for accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use will not infringe privately-owned rights. Mention of any product, its manufacturer, or suppliers shall not, nor is it intended to, imply approval, disapproval, or fitness for any particular use. A royalty-free, nonexclusive right to use and disseminate same for any purpose whatsoever, is expressly reserved to the United States and the University.

# BOOTSTRAP CONFIDENCE INTERVALS

*Robert J. Tibshirani*  
*Department of Statistics*  
*Stanford University*

and  
*Stanford Linear Accelerator Center*

## Abstract

We describe the various techniques that have been proposed for constructing non-parametric confidence intervals using the bootstrap. These include bootstrap pivotal intervals, percentile and bias-corrected percentile intervals, and non-parametric tilting intervals. These methods are small sample improvements over the usual  $\pm t\hat{\sigma}$  intervals. We discuss them in detail, outlining the underlying assumptions in each case. We show how the non-parametric tilting interval can be viewed as an extension of a bootstrap pivotal interval, and suggest a number of generalizations. Finally, the various intervals are compared in a small simulation study.

Work supported by the Department of Energy under contracts DE-AC03-76SF00515 and DE-AT03-81-ER10843, and by the Office of Naval Research under contracts ONR N00014-81-K-0340 and N00014-83-K-0472 and by the U.S. Army Research Office under contract DAAG29-82-K-0056.



# BOOTSTRAP CONFIDENCE INTERVALS

Robert J. Tibshirani

## 1. Introduction.

Recently, a number of techniques have been proposed for constructing confidence intervals using the bootstrap (see Efron(1981) and Schenker(1983)). These techniques are non-parametric in nature, and are designed to work well over a wide variety of situations. Because they are based on the bootstrap, they can be used in situations in which the "parameter" is an extremely complex functional of the distribution and an exact analysis would be impossible.

In this paper, we describe the various bootstrap methods that have been suggested for constructing confidence intervals and compare them in a few examples. The paper is largely expository, although some new ideas are presented for the construction of intervals through non-parametric tilting (Sections 7 and 8). Unfortunately, we offer little in the way of practical advice for choosing among the various intervals. Research in this area is greatly needed.

## 2. The Problem and Some Notation.

We observe  $x_1, \dots, x_n$  assumed to be realizations of random variables  $X_1, \dots, X_n \sim \text{i.i.d } F$ . The distribution  $F$  is unknown and the problem is to construct a confidence interval for the parameter  $\theta = \theta(F)$ . By a confidence interval, we mean lower and upper points  $L = L(x_1, \dots, x_n)$  and  $U = U(x_1, \dots, x_n)$  such that  $P(L \leq \theta \leq U) = 1 - 2\alpha$ , where  $P(\cdot)$  denotes probability under the true distribution  $F$ . Since the intervals are to be non-parametric, we would ideally require that this hold for all  $F$ . None of the intervals described here claim to satisfy this, although the non-parametric tilting interval seems to come closest. We will confine our discussion to central intervals, i.e. intervals  $(L, U)$  such that  $P(\theta \leq L) = P(\theta \geq U) = \alpha$ . Non-central intervals can be obtained through obvious modification.

Given  $X_1, X_2, \dots, X_n$ , ( $X_i$  can be a scalar or vector random variable), we estimate  $\theta$  by  $\hat{\theta} = \theta(F_n^X)$  where  $F_n^X$  is the empirical distribution function of  $X_1, \dots, X_n$ . The observed value of  $\hat{\theta}$  is  $\hat{\theta}_{obs} = \theta(F_n)$  where  $F_n$  is the empirical distribution function of  $x_1, \dots, x_n$ .

We let  $W$  be a random vector with  $W_i \geq 0$ ,  $\sum_1^n W_i = 1$  and  $w$  be a realization of  $W$ . Let  $F_n(w)$  be the distribution putting mass  $w_i$  on  $x_i$ ,  $i = 1, 2, \dots, n$ . Many of the techniques will utilize "bootstrap sampling"—that is, sampling from  $x_1, x_2, \dots, x_n$  with replacement. This is equivalent to sampling  $W$  from the rescaled multinomial  $Mult(n, w^0)/n$ , where  $w^0 = (1/n, 1/n, \dots, 1/n)$ . We'll use  $\sim$  to indicate bootstrap sampling and a bootstrap value obtained in this way will be denoted by  $\hat{\theta}^* = \hat{\theta}(F_n(w))$ . We'll refer to a bootstrap sample either by its weight vector  $w$ , or by  $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ . Finally,  $\hat{B}$  will denote the empirical

## 2 Section 3: Overview

distribution function of  $\hat{\theta}^*$  under  $\mathcal{L}$  ("the bootstrap distribution").

### 3. Overview.

Frequentist confidence intervals are based on a test function, say  $t(\mathbf{X}, \theta_1)$ , appropriate for testing  $H : \theta = \theta_1$ . The interval is constructed as follows. For each trial value  $\theta_1$ , we include  $\theta_1$  in our confidence interval if we would accept  $H$  in a  $1 - 2\alpha$  size test based on  $t(\mathbf{X}, \theta_1)$ . This procedure requires knowledge of the distribution of  $t(\mathbf{X}, \theta_1)$  for each  $\theta_1$ . Usually, a simplifying assumption is made— that  $t(\mathbf{X}, \theta_1)$  is *pivotal*, that is, has a distribution not depending on  $\theta_1$ . With this assumption, it is not necessary to consider each trial value  $\theta_1$  separately. We assume some parametric distribution for  $t(\mathbf{X}, \theta_1)$ , then invert the pivotal to yield the confidence interval. A simple example is  $X_1, X_2, \dots, X_n \sim \mathcal{N}(\theta, 1)$ . Then a confidence interval for  $\theta$  is found by inverting the pivotal  $\bar{X} - \theta$ , whose distribution is  $\mathcal{N}(0, 1/n)$ .

The *Bootstrap Pivotal, Percentile and Bias-Corrected Percentile* intervals (Sections 4, 5, and 6) are non-parametric analogues of parametric pivotal intervals. The pivotal distribution is not assumed known; instead it is estimated non-parametrically using the bootstrap. In Sections 4 and 5 we provide the "recipes" for constructing these intervals and outline the underlying assumptions. In Section 6, we discuss the appropriateness of the various intervals in a few simple problems.

The *Non-Parametric Tilting* interval (Section 7) is more ambitious than the pivotal-based techniques. Instead of assuming the existence of a pivotal, it attempts to estimate the distribution of  $\hat{\theta} - \theta_1$  for each trial value  $\theta_1$ . A confidence interval is then formed consisting of values of  $\theta_1$  for which we would accept the hypothesis  $H : \theta = \theta_1$ .

The *Estimated Pivotal interval* (Section 8) is a compromise between the two approaches. The data is used to suggest an appropriate pivotal quantity, then a bootstrap pivotal interval is constructed.

In Section 9 we compare all the intervals in two numerical examples.

## 4. Bootstrap Pivotal Intervals.

### 4.1. The Simple Pivotal

We assume that  $\hat{\theta} - \theta$  is a pivotal quantity, that is

$$\hat{\theta} - \theta \sim H \quad (A1)$$

where  $H$  is a distribution not involving  $\theta$ , and also that approximately

$$\hat{\theta}^* - \hat{\theta}_{obs} \stackrel{\sim}{\sim} H \quad (A2)$$

Assumption (A2) is based on the premise that if  $F_n$  is close to  $F$ , the bootstrap distribution of  $\hat{\theta}^* - \hat{\theta}_{obs}$  will be close to that of  $\hat{\theta} - \theta$ , as long as  $\theta(\cdot)$  is a reasonably smooth functional. Of course, if  $H$  is a continuous distribution, then (A2) is at best an approximation, since the bootstrap distribution is necessarily discrete. The intervals described in this section and the next section all use this kind of bootstrap approximation. To simplify the notation, *we will ignore the fact that it is only an approximation.*

Under (A1) and (A2), we have  $1 - 2\alpha = P(H^{-1}(\alpha) < \hat{\theta} - \theta < H^{-1}(1 - \alpha)) = P(\hat{\theta} - H^{-1}(\alpha) < \theta < \hat{\theta} - H^{-1}(1 - \alpha))$ .

Substituting  $\hat{\theta}_{obs}$  for  $\hat{\theta}$  and noting that  $H^{-1}(\cdot) = \hat{B}^{-1}(\cdot) - \hat{\theta}_{obs}$ , we obtain the *Bootstrap Pivotal* interval:

$$\theta \in (2\hat{\theta}_{obs} - \hat{B}^{-1}(1 - \alpha), 2\hat{\theta}_{obs} - \hat{B}^{-1}(\alpha)) \quad (1)$$

### 4.2. Other Pivotal

The bootstrap pivotal interval can be based on an arbitrary pivotal  $t(X, \theta)$ , as long as it is monotone in  $\theta$ . We assume  $t(X, \theta) \sim H$ ,  $t(X^*, \hat{\theta}_{obs}) \stackrel{\sim}{\sim} H$ , where  $t(X, \theta)$  is monotone decreasing in  $\theta$ . Inverting the pivot as above, we obtain

$$\theta \in (t_2^{-1}(H^{-1}(1 - \alpha)), t_2^{-1}(H^{-1}(\alpha))) \quad (2)$$

where  $t_2^{-1}(\cdot)$  = inverse of  $t(\cdot, \cdot)$  with respect to the second argument.

The bootstrap pivotal interval is used by Efron (1981) in the form of a "bootstrap  $t$ " and by Schenker (1983), who calls it the "substitution method". We have introduced the obvious name "bootstrap pivotal interval" here.

## 4 Section 5: Percentile Intervals

### 4.3. The Role of Nuisance Parameters

We can think of an arbitrary distribution  $G$  as consisting of two parts, say  $G = (\theta, \eta)$ , where  $\theta = \theta(G)$  is the parameter of interest and  $\eta = \eta(G)$  is a vector of nuisance parameters, possibly infinite dimensional. The true distribution can be written as  $F = (\theta_{true}, \eta_{true})$ . With this decomposition, we can say more clearly the meaning of the statement " $t(X, \theta) \sim H$ ,  $H$  not involving  $\theta$ ". What we're really assuming is that  $F$  is a member of some family of distributions  $\mathcal{F}$  existing in the space of possible distributions. The members of  $\mathcal{F}$  correspond to different  $\theta$  values and are characterized by the property  $t(X, \theta) \sim H$ . Because of this pivotal assumption, we don't have to know the structure of (or estimate) the entire family  $\mathcal{F}$ . Only a single member of  $\mathcal{F}$  need be estimated. The empirical distribution function  $F_n$  estimates that member (i.e.  $(F_n = (\hat{\theta}_{obs}, \hat{\eta}_{obs}))$ ), and from this we obtain the distribution  $H$ . By construction, the interval will have correct coverage for  $F \in \mathcal{F}$ .

A family like  $\mathcal{F}$  also underlies the percentile and bias-corrected percentile intervals (discussed next). The non-parametric tilting interval (Section 7) doesn't make a pivotal assumption, and essentially tries to estimate the entire family.

## 5. Percentile Intervals.

### 5.1. Uncorrected Intervals

Here we assume A1 and A2, and further that

$$H \text{ is symmetric around } 0 \quad (A3)$$

In this case, the pivotal interval (1) becomes:

$$\theta \in (\hat{B}^{-1}(\alpha), \hat{B}^{-1}(1 - \alpha)) \quad (3)$$

Efron calls this the *Percentile Interval* since it uses the percentiles of  $\hat{\theta}^*$  as "percentiles" of  $\theta$ .

### 5.2. Generalization of the Percentile Interval

If a symmetric pivotal exists on some other scale, i.e.

$$g(\hat{\theta}) - g(\theta) \sim H \quad (A4)$$

and

$$g(\hat{\theta}^*) - g(\hat{\theta}_{obs}) \sim H \quad (A5)$$

with  $H$  symmetric around 0 and  $g(\cdot)$  is an unknown, monotone increasing function, then as in (3) we get as an interval for  $g(\theta)$ :

$$g(\theta) \in (\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1 - \alpha)) \quad (4)$$

where  $\hat{G}$  is the distribution function of  $g(\hat{\theta}^*)$ . Transforming back to the  $\theta$  scale gives

$$\theta \in (g^{-1}(\hat{G}^{-1}(\alpha)), g^{-1}(\hat{G}^{-1}(1 - \alpha))) \quad (5)$$

or

$$\theta \in (\hat{B}^{-1}(\alpha), \hat{B}^{-1}(1 - \alpha)) \quad (6)$$

which is again the percentile interval. Thus the percentile interval has the correct coverage if a symmetric pivotal exists on any scale. Conveniently, we don't have to know  $g(\cdot)$  because the resultant interval doesn't depend on  $g(\cdot)$ .

There is a simple connection between the bootstrap pivotal interval based on  $\hat{\theta} - \theta$  and the percentile interval. Writing  $(2\hat{\theta}_{obs} - \hat{B}^{-1}(1 - \alpha), 2\hat{\theta}_{obs} - \hat{B}^{-1}(\alpha))$  as  $(\hat{\theta}_{obs} - [\hat{B}^{-1}(1 - \alpha) - \hat{\theta}_{obs}], \hat{\theta}_{obs} + [\hat{\theta}_{obs} - \hat{B}^{-1}(\alpha)])$ , we see that the percentile interval is the bootstrap pivotal interval reflected about the point  $\hat{\theta}_{obs}$ .

### 5.3. Bias-Corrected Percentile Intervals

#### 5.3.1 Normal Correction.

If the distribution  $H$  in A4 and A5 is symmetric around  $u \neq 0$ , the percentile interval will be biased and will not have the correct coverage. This would occur as a result of bias in the estimator  $\hat{\theta}$ . It turns out that if we are willing to assume a parametric form for  $H$ , then  $u$  can be estimated and a corrected interval can be derived. As was the case for the percentile interval, the corrected interval will not depend on the transformation  $g(\cdot)$ .

Since  $P(g(\hat{\theta}^*) < g(\hat{\theta}_{obs})) = P(\hat{\theta}^* < \hat{\theta}_{obs})$ , we can use the latter to estimate the bias. Using this correction, we then match the distributions of  $g(\hat{\theta}) - g(\theta)$  and  $g(\hat{\theta}^*) - g(\hat{\theta}_{obs})$  on the  $g(\cdot)$  scale, then transform back to the  $\theta$  scale.

As an example, suppose we choose  $H = \mathcal{N}(u, 1)$ . Then

$$g(\theta) - g(\hat{\theta}) \sim \mathcal{N}(0, 1) - u \quad (7)$$

and

$$g(\hat{\theta}^*) - g(\hat{\theta}_{obs}) \sim \mathcal{N}(0, 1) + u \quad (8)$$



## 6 Section 5: Percentile Intervals

We can solve for  $u$  by noting that  $P(g(\hat{\theta}^*) \leq g(\hat{\theta}_{obs})) = \Phi(-u) = \hat{G}(g(\hat{\theta}_{obs})) = \hat{B}(\hat{\theta}_{obs})$  so that  $b \equiv u = -\Phi^{-1}(\hat{B}(\hat{\theta}_{obs}))$ . ( $\Phi(\cdot)$  denotes the cumulative distribution function of  $\mathcal{N}(0, 1)$ ) Now from (7)

$$P(g(\theta) - g(\hat{\theta}) < g(t) - g(\hat{\theta}_{obs})) = \Phi(g(t) - g(\hat{\theta}_{obs}) + b) \quad (9)$$

and from (8) we obtain

$$\hat{B}(t) = P(\hat{\theta}^* \leq t) = P(g(\hat{\theta}^*) \leq g(t)) = \Phi(g(t) - g(\hat{\theta}_{obs}) - b) \quad (10)$$

Solving for  $g(t) - g(\hat{\theta}_{obs})$  in (10) and substituting into (9) we have

$$P(g(\theta) - g(\hat{\theta}) < g(t) - g(\hat{\theta}_{obs})) = \Phi(\Phi^{-1}(\hat{B}(t)) + 2b) \quad (11)$$

Finally, to get a  $1 - 2\alpha$  percent confidence interval, we set the right side of (11) equal to  $\alpha$  and  $1 - \alpha$ , and solve for  $t$  to obtain

$$\theta \in (\hat{B}^{-1}(\Phi(z_\alpha - 2b)), \hat{B}^{-1}(\Phi(z_{1-\alpha} - 2b))) \quad (12)$$

where  $z_p$  denotes the  $p$ th quantile of  $\Phi$ . Interval (12) is called the *Bias-Corrected Percentile Interval*. The parametric assumption  $\mathcal{N}(u, 1)$  turns out to be not as restrictive as it appears. If we instead let  $H = \mathcal{N}(u, \sigma^2)$ , with  $\sigma^2$  unknown, and repeat the above derivation, we get  $b = \frac{u}{\sigma} = -\Phi^{-1}(\hat{B}(\hat{\theta}_{obs}))$  and we obtain the same interval (12).

Note then when  $b = 0$ , the bias-corrected percentile interval reduces to the percentile interval. Hence we can think of the bias-corrected interval as a “fine-tuning” of the percentile interval.

### 5.3.2 Other Symmetric Location Scale Families.

In the bias-corrected interval above, we can just as well assume that  $H$  is some other symmetric, location scale family, say  $H(x | u, \sigma) = H_0(\frac{x-u}{\sigma})$ . This gives the bias-corrected interval

$$\theta \in (\hat{B}^{-1}(H_0(h_\alpha - 2b)), \hat{B}^{-1}(H_0(h_{1-\alpha} - 2b))) \quad (13)$$

where  $b = -H_0^{-1}(\hat{B}(\hat{\theta}_{obs}))$  and  $h_p$  denotes the  $p$ th quantile of  $H_0$ .

A natural question to ask is: how much difference does the choice of  $H_0$  make? Natural candidates to compare with the normal are symmetric, long tailed distributions. Benjamini (1983) provides an appealing definition of long-tailedness. Suppose  $F$  and  $G$  are both symmetric about the origin. Then  $G$  is said to be *stretched* (or long tailed) compared to  $F$  if  $G^{-1}(p)/F^{-1}(p)$  is an increasing function of  $p$ , for  $1/2 < p < 1$ . This definition reflects the intuitive meaning of long-tailedness, that the quantiles of  $G$  are “farther out” than those of  $F$ . Under this definition, distributions like the  $t$ , logistic and cauchy are stretched with respect to the normal, as we would expect.

Now suppose  $H_0$  is stretched with respect to  $\Phi$ . Assume  $\hat{B}(\hat{\theta}_{obs}) = q > .5$ , so that  $\hat{\theta}_{obs}$  is biased upward, and  $b = -\Phi^{-1}(\hat{B}(\hat{\theta})) < 0$ . Then the bias correction under  $H_0$  will be in the same direction as the bias-correction under  $\Phi$ , but will be smaller. The proof of this fact is easily derived from Benjamini's definition above. Denoting, as before, the  $p$ th quantiles of  $\Phi$  and  $H_0$  by  $z_p$  and  $h_p$  respectively, we note that  $H_0(h_\alpha + 2h_q) > \alpha$ . Hence

$$\frac{\Phi^{-1}(H_0(h_\alpha + 2h_q))}{H_0^{-1}(H_0(h_\alpha + 2h_q))} > \frac{\Phi^{-1}(\alpha)}{H_0^{-1}(\alpha)} = \frac{z_\alpha}{h_\alpha} \quad (14)$$

This implies  $\Phi^{-1}(H_0(h_\alpha + 2h_q)) < z_\alpha + 2z_\alpha(h_q/h_\alpha) < z_\alpha + 2z_q$ . Thus  $\Phi(z_\alpha + 2z_q) > H_0(h_\alpha + 2h_q) > \alpha$ .

A similar argument shows that if  $q < .5$ , then  $\Phi(z_\alpha + 2z_q) < H_0(h_\alpha + 2h_q) < \alpha$ , and the corresponding results hold for the upper quantile. The above proof requires that  $h_\alpha + 2h_q < 0$ . This will be the case unless the bias in  $\hat{\theta}_{obs}$  is so large that  $q$  is near  $1 - \alpha$ .

The numbers in Table 1 show the amount of bias correction (that is  $(H_0(h_\alpha + 2h_q), H_0(h_{1-\alpha} + 2h_q))$  for the normal, logistic and the cauchy distributions, when  $\alpha = .05$ .

Table 1

q	Normal	Logistic	Cauchy
.40	(.015, .869)	(.023, .884)	(.045, .944)
.45	(.027, .916)	(.034, .927)	(.050, .950)
.55	(.084, .973)	(.073, .966)	(.050, .950)
.60	(.131, .985)	(.106, .977)	(.056, .955)

The choice of a *symmetric* pivotal distribution appears to make little difference. The effect of an *asymmetric* pivotal distribution, however, can be large, as Example 1 will show.

### 5.3.3 Another Justification for the Bias-Corrected Interval

In place of A4 and A5, we could assume

$$h(\hat{\theta} - \theta) \sim H \quad (A6)$$

and

$$h(\hat{\theta}^* - \hat{\theta}_{obs}) \sim H \quad (A7)$$

with  $H$  symmetric, and  $h$  increasing and anti-symmetric ( $h(-x) = -h(x)$ ). Letting  $H$  be a location-scale family, we again obtain the bias-corrected percentile interval (13). When  $H$  is symmetric around 0,  $\hat{\theta} - \theta$  is symmetric around 0 and the interval reduces to the percentile interval.

Finally, we could replace  $h(\hat{\theta} - \theta)$  and  $h(\hat{\theta}^* - \hat{\theta}_{obs})$  by  $h(\hat{\theta}/\theta)$  and  $h(\hat{\theta}^*/\hat{\theta}_{obs})$  respectively, with  $h(1/x) = -h(x)$ , and again the bias-corrected interval emerges.

## 6. Comparison Between the Bootstrap Pivotal and Percentile Intervals.

The bootstrap pivotal and percentile intervals differ in their assumptions. In constructing the bootstrap pivotal interval, we had to specify the exact form of the pivotal but we assumed nothing about its distribution. On the other hand, in building the percentile interval, knowledge of the exact form of the pivotal was not necessary but we did require that its distribution be symmetric around 0. For the bias-corrected percentile interval, we weakened that assumption to one of symmetry around any point, but we paid a price: it was necessary to specify a distribution for the pivotal.

Which of these intervals is better depends on the problem. It is helpful to look a few simple examples. In each case, the data are assumed to be gaussian.

- *The Mean:*  $\theta = E(X)$ , variance known. The bootstrap pivotal interval based on  $\hat{\theta} - \theta$  and the percentile interval will give very similar results, and both will have approximately the right coverage.
- *The Correlation Coefficient:*  $X = (Y, Z)$  and  $\theta = E(Y - E(Y))(Z - E(Z)) / \{E(Y - E(Y))^2 E(Z - E(Z))^2\}^{1/2}$ . The random variable  $\tanh^{-1}\hat{\theta} - \tanh^{-1}\theta$  is approximately  $N(\theta/(2(n-3)), 1/(n-3))$ . Hence the bootstrap pivotal interval based on  $t(\hat{\theta}, \theta) = \tanh^{-1}\hat{\theta} - \tanh^{-1}\theta$  and the bias-corrected percentile interval (using the normal family) both should work well. The uncorrected percentile interval will be biased.
- *The Variance:*  $\theta = E(X - E(X))^2$ . The random variable  $\hat{\theta}/\theta$  is  $\chi_{n-1}^2$ , hence the bootstrap pivotal based on  $t(\hat{\theta}, \theta) = \log \hat{\theta} - \log \theta$  will have approximately the right coverage. The distribution  $\log \chi^2$  is not symmetric, however, so the percentile intervals may not work well (see Example 1). It is clear that a transformation to a symmetric pivotal doesn't exist here since such a transformation must also remove the dependence of the variance on  $\theta$ . A simple delta method calculation shows that only  $g(\hat{\theta}) = \log \hat{\theta}$  achieves this.

The above examples represent some of the problems that are well understood. In most situations, however, matters are much more difficult. To construct a bootstrap pivotal interval, we first need to specify a quantity  $t(X, \theta)$  that is approximately pivotal. This alone is a difficult task unless we know something about the underlying distribution. Now suppose we are able to specify a pivotal  $t(X, \theta)$ . Then if  $t(X, \theta) \sim H$  and  $t(X^*, \hat{\theta}_{obs}) \sim H$ , the resulting interval will have the correct coverage. In some problems, however, the bootstrap distribution of  $t(X^*, \hat{\theta}_{obs})$  can be a poor approximation to  $H$ . One such example is the following. Consider the situation  $X_1, X_2, \dots, X_{15} \sim e^{-1-x}$  for  $x \geq -1$ . The bootstrap pivotal interval for  $\theta = E(X)$  based on  $\bar{X} - \theta$  has poor coverage because the distribution of  $\bar{X}^* - \bar{X}_{obs}$  is not a good approximation to the distribution of  $\bar{X} - \theta$ . This is because the high positive correlation between  $\bar{X}$  and the sample standard deviation  $S$  causes underestimation of the scale when  $\bar{x}$  is smaller than

## 10    *Section 6: Comparison Between the Bootstrap Pivotal and Percentile Intervals*

$\theta$  and overestimation of the scale when  $\bar{x}$  is greater than  $\theta$ . Basing the interval on  $(\bar{X} - \theta)/S$  alleviates this problem and the resultant interval has good coverage.

As far as the percentile methods are concerned, an important question is: when does a transformation to a pivotal quantity exist? (Efron (1983) discusses normalizing transformations). If such a transformation existed for a broad class of problems, the percentile (and bias-corrected percentile) intervals would be attractive because they eliminate the problem of having to specify a pivotal.

Unfortunately, we have no solid answers to these questions; instead we move on to a different technique that sidesteps these difficulties.

## 7. Non-Parametric Tilting Intervals.

### 7.1. Definition

As mentioned in the overview, the “*Non-parametric tilting*” interval (Efron (1981)) is fundamentally different than the preceding methods. Instead of assuming the existence of a pivotal, it attempts to estimate the distribution of the test function  $\hat{\theta} - \theta_1$  for each value of  $\theta_1$ . Then, as in classical testing theory, a confidence interval is formed consisting of values of  $\theta_1$  for which we would accept the observed data in a test of  $H : \theta = \theta_1$ .

We begin by assuming that the true distribution  $F$  is supported only on  $x_1, \dots, x_n$ . (We’ll see later why this is necessary). Let  $\mathbf{w} = (w_1, \dots, w_n)$ , with  $w_i \geq 0$  and  $\sum_1^n w_i = 1$ . Recall that  $F_n(\mathbf{w})$  is the distribution putting mass  $w_i$  on  $x_i$ ,  $i = 1, 2, \dots, n$  and  $\hat{\theta}(\mathbf{w}) \equiv \theta(F_n(\mathbf{w}))$ . Then the true distribution  $F$  can be represented by  $\mathbf{w}_{true}$ , (i.e.  $F = F_n(\mathbf{w}_{true})$ ). Let  $P_{\mathbf{w}}$  denote probability under  $Mult(n, \mathbf{w})/n$ , and denote the observed distribution by  $\mathbf{w}^0 = (\frac{1}{n}, \dots, \frac{1}{n})$ . Then the *Non-parametric tilting interval* is  $(\theta_{low}, \theta_{up})$  given by

$$\theta_{low} \equiv \inf\{\theta_1 : P_{\mathbf{w}_{\theta_1}}(\hat{\theta}(W) \geq \hat{\theta}_{obs}) \geq \alpha\} \quad (15)$$

and

$$\theta_{up} \equiv \sup\{\theta_1 : P_{\mathbf{w}_{\theta_1}}(\hat{\theta}(W) \leq \hat{\theta}_{obs}) \geq \alpha\} \quad (16)$$

where  $\mathbf{w}_{\theta_1}$  minimizes  $D(\mathbf{w}, \mathbf{w}^0)$  subject to  $\hat{\theta}(\mathbf{w}) = \theta_1$ , and  $D(\mathbf{w}, \mathbf{w}^0)$  is the (backward) *Kullback-Leibler distance* between  $\mathbf{w}$  and  $\mathbf{w}^0$ :

$$D(\mathbf{w}, \mathbf{w}^0) \equiv \sum_1^n w_i \log(nw_i) \quad (17)$$

In words,  $\theta_{low}$  is the smallest parameter value  $\theta_1$  such that if the data came from a distribution with parameter  $\theta_1$ , there is probability at least  $\alpha$  of observing  $\hat{\theta}_{obs}$ . Similarly for  $\theta_{up}$ . Given any value  $\theta_1$ , there may be more than one distribution  $\mathbf{w}$  with  $\hat{\theta}(\mathbf{w}) = \theta_1$ , so we use our “best guess”—the closest distribution in Kullback-Leibler distance to our data, subject to  $\hat{\theta}(\mathbf{w}) = \theta_1$ .

From the definition of the non-parametric tilting interval, it certainly isn’t clear why it should have the correct coverage. We investigate this in Section 7.5. We will first discuss how  $(\theta_{low}, \theta_{up})$  can be computed.



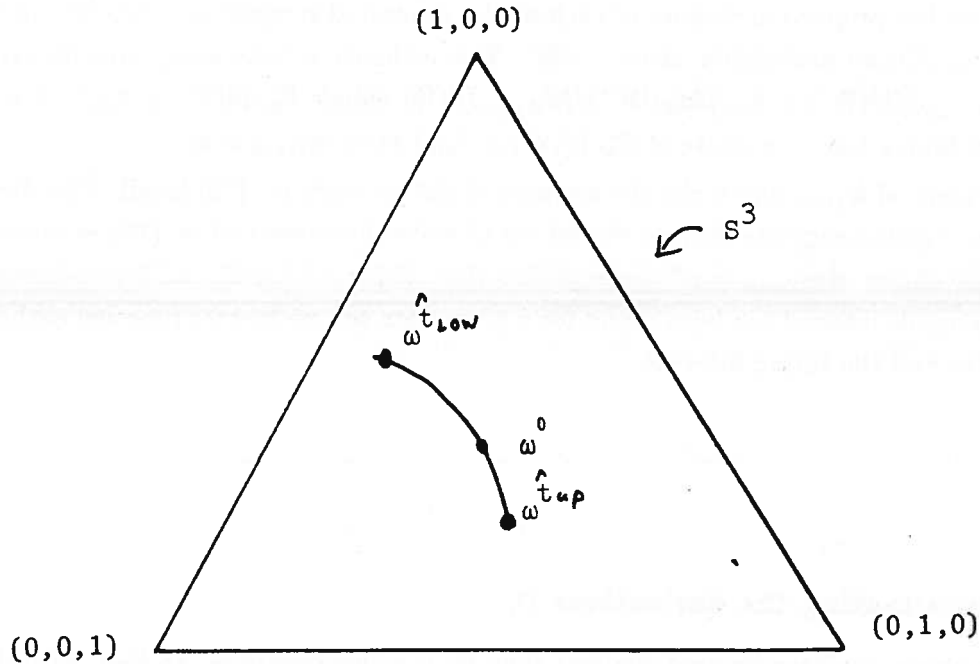
### 7.2. Finding $(\theta_{low}, \theta_{up})$

In order to compute  $(\theta_{low}, \theta_{up})$ , we need to find the distribution  $\mathbf{w}$  that minimizes  $D(\mathbf{w}, \mathbf{w}^0)$  subject to  $\theta(\mathbf{w}) = \theta_1$ . Minimization by the method of lagrange multipliers yields weights of the form

$$w_i^t = \frac{e^{D_i t}}{\sum_1^n e^{D_i t}} \quad (18)$$

where  $D_i = \frac{d\hat{\theta}(\mathbf{w})}{dw_i}$  evaluated at  $\mathbf{w}^t$ . Note that the observed distribution  $\mathbf{w}^0$  corresponds to  $t = 0$ . The one dimensional family (18) contains the closest distributions to  $\mathbf{w}^0$  for each fixed value of  $\hat{\theta}(\mathbf{w})$ . It is illustrated in Figure 1.

Figure 1



The triangle represents the simplex  $S^n = \{\mathbf{w} : w_i \geq 0, \sum_1^n w_i = 1\}$  for  $n = 3$ . The statistic  $\hat{\theta} = \theta(F_n(\mathbf{w}))$  can be thought of a surface defined over the simplex, and the curve passing through the sample point  $\mathbf{w}^0$  is the one-dimensional family defined by (18). Consider any distribution (say  $\mathbf{w}'$ ) in this family. Suppose  $\theta(\mathbf{w}') = \theta'$ . Then  $\mathbf{w}'$  is the closest distribution to  $\mathbf{w}^0$  (i.e. minimizes  $D(\mathbf{w}, \mathbf{w}^0)$ ) having  $\theta$  value  $\theta'$ .

To find  $\theta_{low}$  and  $\theta_{up}$ , it is necessary to compute  $P_{\mathbf{w}^t}(\hat{\theta}(W) \geq \hat{\theta}_{obs})$  and  $P_{\mathbf{w}^t}(\hat{\theta}(W) \leq \hat{\theta}_{obs})$  for many values of  $t$ . When  $\hat{\theta}(W)$  is the sample mean, the distribution of  $\hat{\theta}(W)$  under any  $\mathbf{w}^t$  turns out to be a simple exponential tilt of the distribution under  $\mathbf{w}^0$ . As a result, it is only necessary to simulate the distribution of  $\hat{\theta}(W)$  under  $\mathbf{w}^0$ ; the remaining distributions can then be derived analytically. This is the reason for the name "non-parametric tilting interval" (see Efron (1981) for details).

For non-linear statistics, the tilting property no longer holds. Hence to determine  $\theta_{low}$ , we need to search along  $\mathbf{w}^t$  until  $P_{\mathbf{w}^t}(\hat{\theta}(W) \geq \hat{\theta}_{obs})$  is exactly equal to  $\alpha$ . (and similarly for  $\theta_{up}$ ). (We have labelled these distributions  $\mathbf{w}^{low}$  and  $\mathbf{w}^{up}$  in Figure 1). Now for each  $t$ , direct computation of  $P_{\mathbf{w}^t}(\hat{\theta}(W) \geq \hat{\theta}_{obs})$  requires a monte carlo approximation, and this would make the search procedure far too costly. Instead, we can generate  $B$  samples from a fixed distribution  $\mathbf{w}_{fixed}$ , and use as our monte carlo estimate:

$$\hat{P}_{\mathbf{w}^t}(\hat{\theta}(W) \geq \hat{\theta}_{obs}) = \frac{1}{B} \sum_1^n I(\hat{\theta}(W^i) \geq \hat{\theta}_{obs}) \frac{M_{\mathbf{w}^t}(W^i)}{M_{\mathbf{w}_{fixed}}(W^i)} \quad (19)$$

where  $W^i$  are the proportion vectors of each of the generated samples and  $M_{\mathbf{w}}(W)$  indicates the multinomial( $n, \mathbf{w}$ ) probability of vector  $W$ . This estimate is unbiased since its expected value is  $E_{\mathbf{w}_{fixed}}\{I(\hat{\theta}(W^i) \geq \hat{\theta}_{obs}) M_{\mathbf{w}^t}(W^i) / M_{\mathbf{w}_{fixed}}(W^i)\}$  equals  $P_{\mathbf{w}^t}(\hat{\theta}(W) \geq \hat{\theta}_{obs})$ . It reduces to the usual Monte Carlo estimate of  $P_{\mathbf{w}^t}(\hat{\theta}(W) \geq \hat{\theta}_{obs})$  when  $\mathbf{w}_{fixed} = \mathbf{w}^t$ .

Wise choice of  $\mathbf{w}_{fixed}$  can make the variance of the estimate in (19) small. The theory of importance sampling suggests that we should try to make the summand in (19) as constant as possible. The choice of  $\mathbf{w}_{fixed} = \mathbf{w}^0$  accomplishes this. Using  $\mathbf{w}^0$  has the additional advantage that if a percentile interval has been found for a given data set, no further function evaluations are needed to find the tilting interval.

### 7.3. Approximating the derivatives $D_i$

In the above, we have ignored the fact that  $D_i$  is a function of  $\mathbf{w}$ , so that (18) defines  $\mathbf{w}^t$  only implicitly. As an approximation to  $D_i$ , we use  $D_i^0 = \frac{d\hat{\theta}(\mathbf{w})}{d\mathbf{w}_i}$  evaluated at  $\mathbf{w}^0$ . If  $\theta(\mathbf{w})$  is too complex to make calculation of  $D_i^0$  possible, we can use the jackknife estimate of the directional derivative  $U_i$  given by

$$\hat{U}_i = (n-1)[\hat{\theta}(F_n(\mathbf{w}_{-i})) - \hat{\theta}_{obs}] \quad (20)$$

where  $\mathbf{w}_{-i} = (1/(n-1), 1/(n-1), \dots, 1/(n-1), 0, 1/(n-1), \dots, 1/(n-1))$ , zero being in the  $i$ th place. Since  $D_i^0 = U_i + \text{constant}$ , (see Efron(1982) pg 37-41 for details),  $\hat{U}_i$  can be used in place of  $D_i^0$  in formula (18).

#### 7.4. The Non-parametric Tilting Interval as a Generalization of the Bootstrap Pivotal Interval

The interval  $(\theta_{low}, \theta_{up})$  can be derived as a generalization of a bootstrap pivotal interval based on  $\hat{\theta} - \theta$ . In the bootstrap pivotal case, we proceeded as follows. We began by assuming that  $\hat{\theta} - \theta \sim H$ ,  $H$  not involving  $\theta$ . We then estimated  $H$  by assuming that the bootstrap version of this,  $\hat{\theta}^* - \hat{\theta}_{obs}$ , also had the distribution  $H$ . Using this bootstrap distribution, the pivotal was inverted to yield the interval. Note that the bootstrap approximation consisted solely of estimating the unknown  $F$  by  $F_n$ , the unrestricted maximum likelihood estimator of  $F$ .

In the present case, we don't assume that  $\hat{\theta}(W) - \theta$  is pivotal, and hence we'll write  $\hat{\theta}(W) - \theta \sim H_\theta$  to indicate that  $H$  depends on  $\theta$ . In order to construct the interval, we need to estimate  $H_{\theta_1}$  for each trial value  $\theta_1$ . To do so, we first compute the restricted maximum likelihood estimator of  $F$ , subject to  $\theta(F) = \theta_1$ . Since we are assuming that  $F$  has support only on  $x_1, x_2, \dots, x_n$ , this restricted m.l.e. corresponds to a set of weights, say  $w_{\theta_1}$ . Proceeding as in the pivotal case, we assume that under  $w_{\theta_1}$ ,  $\hat{\theta}(W) - \theta_1$  also has distribution  $H_{\theta_1}$ . Now to obtain the confidence interval, we can't just "invert the pivot" since  $H_{\theta_1}$  depends on  $\theta_1$ . Instead, we consider each  $\theta_1$  separately, including  $\theta_1$  in the confidence interval if  $P_{w_{\theta_1}}(\hat{\theta}(W) - \theta_1 \geq \hat{\theta}_{obs} - \theta_1) \geq \alpha$  and  $P_{w_{\theta_1}}(\hat{\theta}(W) - \theta_1 \leq \hat{\theta}_{obs} - \theta_1) \geq \alpha$ . Finally, equivalence between maximizing the multinomial likelihood  $L = \prod_1^n w_i$  and minimizing the Kullback-Leibler distance almost yields (15) and (16). We say "almost" because maximizing  $L$  subject to  $\hat{\theta}(w) = \theta_1$  is equivalent to minimizing the "forward" Kullback-Leibler distance  $D(w^0, w) = \sum \frac{1}{n} \log \frac{1}{nw_i}$  subject to  $\hat{\theta}(w) = \theta_1$ . Strictly for computational reasons \* we instead use the "backward" Kullback-Leibler distance  $D(w, w^0)$ .

Note that if we don't restrict the support of  $F$  to  $x_1, x_2, \dots, x_n$ , the maximum likelihood estimator of  $F$  subject to  $\theta(F) = \theta_1$  is pathological in many cases, and not of much use in constructing the interval. Alternatively, we can impose smoothness restrictions on  $F$  (see Section 7.6).

Figure 2 illustrates the level surfaces  $C_\theta$  of constant  $\theta$  ( $C_\theta = \{w : \hat{\theta}(w) = \theta\}$ ) together with the various quantities used in deciding whether to include  $\theta$  in  $(\theta_{low}, \theta_{up})$ . (The distribution  $w_{\theta, W}$  appearing in Figure 2 should be ignored for now—it will be important in discussing a variant of this interval in Section 7.6). Figure 2 makes clearer the distinction between the bootstrap pivotal and non-parametric tilting intervals. Both intervals use the distance function  $\hat{\theta}(W) - \theta$  to measure the distance between a point  $W$  and the level surface  $C_\theta$ . The bootstrap pivotal method assumes that  $\hat{\theta}(W) - \theta$  has a distribution not depending on  $\theta$ , i.e. the distribution of  $\hat{\theta}(W) - \theta$  is the same no matter which level surface  $w_{true}$  is on. As a result, we don't have to estimate the distribution of  $W$  for each  $\theta$ , and we need only sample from the

---

\* The forward Kullback-Leibler distance yields weights of the form  $c/(1 + tD_i)$  which we have found difficult to deal with computationally.



non-parametric tilting interval. One such attempt is made in Section 7.7.

### 7.6. Smooth Tilted Intervals

Instead of assuming that  $F$  is supported on  $x_1, x_2, \dots, x_n$ , we could assume instead that  $F$  belongs to a smooth family. A convenient choice for this family is  $F_n(\mathbf{w})$  convolved with a symmetric kernel  $k(\cdot)$ :

$$g_{\mathbf{w}}(y) = \sum_1^n \frac{w_i}{h} k\left(\frac{y - x_i}{h}\right) \quad (21)$$

The window parameter  $h$  can be estimated from the data or more simply, a reasonable fixed value (say  $h = .25$ ) can be used.

The form of the weights that minimize  $D(\mathbf{w}, \mathbf{w}^0) = E_{g_{\mathbf{w}}} \log g_{\mathbf{w}} / g_{\mathbf{w}^0}$  appears quite difficult to obtain for general kernels  $k$ . (Since the weights are to be used as part of a computationally intensive procedure, an explicit expression is required.) All is not lost, however. When the supports of the  $k(\frac{y-x_i}{h})$ 's are disjoint, the minimizing weights are exactly of the form (18), independent of  $k(\cdot)$ . Thus, as an approximation, we can use weights of the form (18). This should be adequate for small  $h$ .

### 7.7. Another Variant of the Non-Parametric Tilting Interval

A variant of the  $(\theta_{low}, \theta_{up})$  interval can be obtained by use of a test function other than  $\hat{\theta}(W) - \theta$ . Recall that  $\mathbf{w}_{\theta}$  is the closest point (in backward Kullback-Leibler distance) on  $C_{\theta}$  to  $\mathbf{w}^0$ . Then a natural measure of the plausibility of a value  $\theta$  is the distance  $D(\mathbf{w}_{\theta}, \mathbf{w}^0)$ . In fact, we will also want to take into account on which "side" of  $C_{\theta}$  the observed distribution  $\mathbf{w}^0$  lies. This can be achieved by attaching a sign to the distance in some consistent fashion. The interval defined below is similar to the one introduced by Efron (1984) for a certain class of parametric problems.

For an arbitrary distribution  $\mathbf{w}$ , let  $\mathbf{w}_{\theta, \mathbf{w}}$  be the closest point in backward Kullback-Leibler distance on  $C_{\theta}$  to  $\mathbf{w}$ . Define the signed (backward) Kullback-Leibler distance by  $SD(\mathbf{w}_{\theta, \mathbf{w}}, \mathbf{w}) = D(\mathbf{w}_{\theta, \mathbf{w}}, \mathbf{w})$  if  $\theta \geq \hat{\theta}(\mathbf{w})$  and  $-D(\mathbf{w}_{\theta, \mathbf{w}}, \mathbf{w})$  otherwise. Now consider a trial value  $\theta_1$ . Then we include  $\theta_1$  in our confidence interval if

$$P_{\mathbf{w}_{\theta_1}, W}(SD(\mathbf{w}_{\theta, W}, W) \geq SD(\mathbf{w}_{\theta_1}, \mathbf{w}^0)) \geq \alpha \quad (22)$$

and

$$P_{\mathbf{w}_{\theta_1}, W}(SD(\mathbf{w}_{\theta, W}, W) \leq SD(\mathbf{w}_{\theta_1}, \mathbf{w}^0)) \geq \alpha \quad (23)$$

Figure 2 again illustrates this, the only difference being that the distance from a point  $W$  to the level surface  $C_{\theta}$  is measured by minimum Kullback-Leibler distance instead of by  $\hat{\theta}(W) - \theta$ .

How would this interval compare to the  $(\theta_{low}, \theta_{up})$  interval? If the level curves are such that  $SD(\mathbf{w}_\theta, \mathbf{W}, \mathbf{w}) = \theta - \hat{\theta}(\mathbf{W})$ , (or some monotone function of  $\theta - \hat{\theta}(\mathbf{W})$ ) then the two methods are identical. This follows simply by the definition of the two intervals. In general, however, they will be different. The parametric analogues of these procedures suggest that the interval based on  $SD(\mathbf{w}_\theta, \mathbf{W}, \mathbf{W})$  might be less sensitive to departures of  $\mathbf{w}_\theta$  from  $\mathbf{w}_{true}$  and to the changing shape of the level surfaces. On the other hand, it appears to be much more difficult to compute. The reason is that for *each bootstrap sample*, the distance  $SD(\mathbf{w}_\theta, \mathbf{W}, \mathbf{W})$  must be computed, and this requires an expensive search. (Only one such search was necessary in computing the non-parametric tilting interval.) Some clever computing tricks might make computation of this interval possible, but this is still under investigation.

## 8. Estimated Pivotal Intervals.

### 8.1. Definition

We propose here a compromise between the pivotal methods and the non-parametric tilting technique. The idea is to use the estimated distributions provided by the tilting technique to estimate the form of a pivotal. This pivotal is then used as the basis for a bootstrap pivotal interval (as described in Section 4).

Assume  $\hat{\theta} - \theta \sim H$  and  $\hat{\theta}^* - \hat{\theta}_{obs} \sim H$  where  $H$  has variance  $v(\theta)$ . Then a one term Taylor expansion shows that  $g(\hat{\theta}) - g(\theta)$  has approximately constant variance, where

$$g(\theta) = \int_c^\theta \frac{1}{v(t)^{\frac{1}{2}}} dt \quad (24)$$

A convenient family for  $v(\theta)$  is  $v_p(\theta) = |\theta|^p$  for  $0 \leq p \leq 2$ . Then  $g_p(\theta) = |\theta|^{1-p/2}$  if  $p \neq 2$  and  $g_2(\theta) = \log |\theta|$ .

### 8.2. Estimation of p

We use the family of closest distributions to  $\mathbf{w}^0$  to estimate  $v(\theta)$  for various values of  $\theta$ . These estimates are obtained using the computational trick of Section 7. A simple linear regression of  $\log \hat{v}(\theta)$  on  $|\theta|$  then provides an estimate of  $p$ .

Given the estimate  $\hat{p}$ , we then utilize the pivotal  $g_{\hat{p}}(\hat{\theta}) - g_{\hat{p}}(\theta)$  in the bootstrap pivotal interval given by (2).



## 9. Some Numerical Examples.

To illustrate the various methods discussed here, we compared them in two problems.

**Example 1. The Variance:**  $\theta = E(X - E(X))^2$ ,  $X \sim \mathcal{N}(0, 1)$ .

As we discussed earlier, we would expect that the bootstrap pivotal interval based on  $\hat{\theta}/\theta$  to perform well, since  $\hat{\theta}/\theta$  is a pivotal, but would expect the percentile and bias-corrected percentile intervals not to do as well.

We performed a small simulation study to investigate this. 1000 samples of size 20 were generated and central 90 percent confidence intervals were constructed for each sample. The results are shown in Table 2. The normal theory interval obtained by pivoting  $\hat{\theta}/\theta$  around the percentage points of  $\chi^2_{19}$  can be thought of as the "correct" interval and had actual coverage close to 90 percent. The bootstrap pivotal interval does fairly well, while the others display too low coverage. The percentile interval is especially poor. The non-parametric tilting and estimated pivotal intervals captured the asymmetry of the normal interval better than the bias-corrected interval, but their coverage was still too low.

It's not surprising that the intervals based on  $\hat{\theta} - \theta$  or  $g(\hat{\theta}) - g(\theta)$  didn't perform well: this particular problem is difficult because the variance of  $\hat{\theta}$  depends on  $\theta$ . Only the bootstrap pivotal interval, which utilizes knowledge of this fact, performs satisfactorily.

Table 2 also displays the results for the "smooth" versions of these intervals. All bootstrapping was done from  $F_n(\mathbf{w}) * .25\mathcal{N}(0, \hat{\theta}_{obs})$  instead of  $F_n(\mathbf{w})$ . For the bootstrap pivotal, percentile and bias-corrected percentile intervals, this meant sampling a data value with replacement from  $x_1, x_2, \dots, x_n$ , then adding .25 times a  $\mathcal{N}(0, \hat{\theta}_{obs})$  random variable. The approximate weights (18) were used for the non-parametric tilting interval. The smoothing improved the coverage of all the bootstrap intervals, but surprisingly, it pulled the bias-corrected and non-parametric tilting intervals away from the normal interval.

Finally, we tried the unsmoothed intervals with sample 50. All the intervals performed quite well except the percentile interval which was biased.

This problem was also studied by Schenker(1983). He obtained results for bootstrap pivotal and percentile intervals; they are in qualitative agreement with those given here.

**Example 2. *The 10 percent Trimmed Mean***

1000 Samples of size 20 were generated from the contaminated normal  $.90N(0,1) + .10N(0,3)$  and central 90 percent confidence intervals were computed. For comparison with the bootstrap interval, we computed the asymptotic interval  $(\hat{\theta} - t_{.95}s_{jk}, \hat{\theta} + t_{.95}s_{jk})$  where  $t_{.95}$  is the 95th percentile of the  $t_{19}$  distribution and  $s_{jk}$  is the jackknife estimate of standard deviation. The results are shown in Table 3. All the intervals had approximately the right coverage; surprisingly, the bias-correction seemed to make the percentile interval worse.

## **10. Closing Remarks.**

We have discussed a number of bootstrap techniques for constructing confidence intervals. All are potentially useful as data-analytic tools because they are non-parametric and can be applied in complex situations. Further work is needed to evaluate and improve these methods. Our current research focusses the non-parametric tilting interval and its variants.

### **Acknowledgements**

The author wishes to thank Bradley Efron, Trevor Hastie, Tim Hesterberg, M. Vernon Johns and Werner Stuetzle for valuable discussions. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

#### REFERENCES

- Benjamini, Y. (1983). Is the t-test really conservative when the parent distribution is long-tailed? *J. Amer. Stat. Assoc.* 78, pp 645-654
- Efron, B. (1981). Non-parametric standard errors and confidence intervals (with discussion). *Can. J. Stat.* Vol 9., No 2. pp 139-172
- Efron, B.(1982) The Jackknife, bootstrap, and other resampling plans. *SIAM Monograph*.
- Efron, B.(1983) Transformation theory: How normal is a family of distributions? Dept. of Statistics, Stanford Univ. Tech Rep. 173.
- Efron, B.(1984) Bootstrap confidence intervals for parametric problems. Dept. of Statistics, Stanford Univ. Tech Rep. 90.
- Schenker, N.(1983) A closer look at bootstrap confidence intervals. Department of Statistics, University of Chicago Tech. Rep. 150

Table 2

## Confidence Intervals for the Sample Variance

Sample Size: 20  
 Monte Carlo size: 1000  
 Number of Bootstraps: 500

	Ave Lower	Ave Upper	Level (Standard Error)
Normal	.632	1.883	9.8% (0.94)
Pivotal	.670	1.860	15.7% (1.15)
Percentile	.496	1.357	23.3% (1.34)
Bias-C Percentile	.595	1.490	20.9% (1.29)
Tilting	.634	1.552	20.2% (1.27)
Estimated Pivotal	.587	1.520	19.3% (1.25)
Pivotal Smoothed	.625	1.752	13.8% (1.09)
Perc Smoothed	.525	1.452	19.3% (1.25)
Bias-C Smoothed	.539	1.472	18.8% (1.24)
Tilting Smoothed	.568	1.527	17.4% (1.20)

Sample Size: 50

Normal	.723	1.404	10.3% (0.96)
Bootstrap Pivotal	.750	1.431	11.7% (1.02)
Percentile	.670	1.272	16.9% (1.19)
Bias-C Percentile	.718	1.335	14.7% (1.12)
Tilting	.738	1.372	13.4% (1.08)
Estimated Pivotal	.718	1.341	13.3% (1.07)

Table 3

## Confidence Intervals for the 10% trimmed mean

Sample size: 20  
 Monte Carlo Size: 1000  
 Number of Bootstraps: 500

	Ave Lower	Ave Upper	Level (Standard Error)
T-interval	-.4427	.4590	10.1% (0.95)
Bootstrap Pivotal	-.4585	.4764	8.3% (0.87)
Percentile	-.4616	.4777	10.7% (0.98)
Bias-C Percentile	-.4825	.4953	8.3% (0.87)
Tilting	-.4510	.4694	11.1% (0.99)